

## Detection of Short-Term Fluctuations in Measurement Systems by Determination of Extreme Values and Range

Erwin Hansert

Max-Planck-Institut für Psychiatrie, Klinisches Institut – Biostatistik, München, Germany

*Dedicated to Prof. Dr. Dr. J. Büttner on the occasion of his 65th birthday*

**Summary:** In the new draft, in preparation, of the guidelines of the German Medical Council (Richtlinien der Bundesärztekammer, RiLiBÄK) for quality management of clinical laboratory measurement systems, particular emphasis is placed on the establishment of a lower and upper limit for each control value in a control cycle. In the present study, an attempt is made to show that this approach alone does not detect even relatively large short-term changes in a measurement system. A substantial improvement is possible with modest outlay, by also examining the difference between the largest and smallest individual control values, i. e. by determining the current range. Appropriate practical limits for this procedure are derived from the theoretical behaviour of extreme values, based initially on the usual assumption of a normal distribution. Furthermore, it is shown that these limits are also meaningful for other assumed appropriate distributions.

### Introduction

The reliable medical evaluation of values from patient samples depends on the constancy (stability) of the measurement systems employed. Such measurements are frequently performed in order to:

- compare a patient value with a reference range
- determine whether a patient value has changed with respect to earlier values from the same patient.

The constancy of a measurement system must be continually monitored with suitable samples (control samples), i. e. samples for which a “(conventional) true value” has been determined, which can be employed as the target value. In the first instance, this constancy or lack of constancy is always expressed statistically or phenomenologically, since it is based on the ongoing measurement of control samples with target values. According to the new draft of the guidelines (1), in each control cycle in which control samples are analysed, the lower and upper limits must be found for:

- individual measurements (deviations from the target value)
- average values (systematic deviations)
- standard deviations (random deviations).

As a rule, this means that one control value per day selected independently of the measurement results is determined. According to l. c. (1), when a control sample is applied for the first time, control of the average value and standard deviation of the first cycle are not

applicable. Also, *during* a control cycle it is only to determine whether individual measurements lie within the stipulated limits.

The first problem is that these limits must all be established on the basis of general considerations. The second problem concerns the effectiveness of the overall principle of the control procedure, which consists of adhering to generally accepted limits. In my opinion the recommended measures need to be supplemented. Experience with the determination of assigned values shows that measurement systems are often subject to undesirable short-term changes. Such changes cannot be detected merely by looking for compliance with generally defined limits (see examples given below).

For this purpose, it is necessary to look for an additional procedure in each control cycle, which is simple to perform and does not entail a significant increase in cost. Of course, this does not affect the implementation of a more thorough management of the control procedure in any individual laboratory. A suggestion for a procedure of this kind is the evaluation of the current range of the individual values of control samples during the control cycle. The mathematical, or rather the numerical basis of such a procedure is presented here. The influence of the distribution of measurement values is also considered; these are usually assumed to have a normal distribution. It is shown that the numerical criteria derived for a normal distribution also apply, without substantial alteration, to the “Q-distributions” (see appendix).

## The Model

The normal distribution is the classical model for describing and evaluating the results of quantitative observations (measurements results in the narrow sense). This has been the case since *C. F. Gauss* showed that the summation of a large number of uniformly negligible random variables leads to a normal distribution (*Gaussian error law*). Normal distributions are uniquely defined by the (theoretical) mean value and the (theoretical) standard deviation. Furthermore, the arithmetic mean and empirical standard deviation carry the greatest weight, i. e. the maximal likelihood. The normal distribution is therefore the obvious model:

- a) when dealing with a typical, i. e. mean value, and
- b) when the measurement results can be completely or at least adequately described by the mean value and standard deviation.

If, however, individual results must be kept within predetermined limits in a measurement series, then the crucial values are not the mean values, but the extreme values that do or do not exceed these limits. If these limits must be set so wide that they are virtually never exceeded for the most of the accepted operational conditions of a measurement system, then they are inadequate for testing stability. On the other hand, the limits must not be set so narrowly that they are exceeded purely at random and with high probability. This means that limits are suitable for testing stability only if they are predetermined for the individual measurement system. They must be chosen so that, in a stable measurement system, they are not exceeded too often purely at random, while displaying adequate sensitivity to changes in the operational performance of the measurement system. For this purpose, we recommend in principle a probability value of 10% for the random exceeding of limits (see also Discussion). We must therefore consider the behaviour of the extreme values from a series of measurements in an individual measurement system. At the same time, in accordance with the guidelines (1), the mean value and standard deviation of the measured control values must also be taken into account.

Such questions have already been adequately investigated for normal distributions (see, e. g. l. c. (2)). However, certain difficulties arise from basing this treatment on the assumption of a normal distribution. Thus normal distributions have an unlimited scale either side of the mean value. This means that the "carrier" of each normal distribution (i. e. those scale values for which the probability density function is greater than 0) is basically the entire axis. Therefore, as the number of measurements increase, it is to be expected that individual values will display ever increasing deviations from the mean. However, a measurement system with this property

would be of no use. Consider, for example, a system for measuring serum glucose concentration, and a control sample with a glucose concentration of 5.56 mmol/l (100 mg/dl). If the distribution of measured values is really normal, then as the number of measurements increases, results are to be expected that deviate from 5.56 mmol/l (100 mg/dl) by any multiple of the analytical standard deviation (imprecision), e. g. by at least six-fold. For an analytical standard deviation of e. g. 0.28 mmol/l (5 mg/dl) results would be above 7.22 mmol/l (130 mg/dl) and below 3.89 mmol/l (70 mg/dl).

The question therefore arises as to whether, on the basis of a normal distribution, unrealistic criteria are derived for the evaluation of individual measurements. For this reason, a second family of distributions was considered, so called here "Q-distributions", and their subfamily of "standardized Q-distributions" (see Appendix). Just like normal distributions, all of these distributions are symmetrical about the mean, but they have a finite carrier, i. e. there are limits to the possible measurement values above and below the mean. In addition, as the carrier is suitably extended, these distributions converge to normal distributions. The acceptability of the following suggestions is supported by the fact that the numerical criteria derived for normal distributions need not be fundamentally altered to accommodate other types of distribution.

## Extreme Values and Range

In the following it must always be borne in mind that there is a difference between the actual measured values and the accompanying random values (the model values), i. e. between the distribution of measured values on the measurement axis and the underlying probability distribution. Here, we are considering only the mathematical model, which consists of a series of random values  $X_1, X_2, \dots, X_n$ . They must be regarded as (stochastically) independent, and therefore model representatives of a series of measurements with no mutual influences. In a stable measurement system, all  $X_i$  have the same probability distribution. If changes occur in the measurement system, this latter assumption is no longer valid, since such changes lead to alterations in the probability distribution, in particular in the properties of extreme values. The extreme values respond more or less quickly and strongly to the inconstancy of the measurement system, depending on its type and magnitude.

Let the measured values of these random quantities be arranged according to their size. This leads to new random quantities known as ranks:

$$(1) \quad X_{(1)} < X_{(2)} < \dots < X_{(n-1)} < X_{(n)}$$

(In the continuous probability distribution considered here, equals signs are unnecessary, i. e. all the measured

**Tab. 1** Some characteristic quantities for the largest measurement value from a random sample of size  $n = 10$  and  $n = 20$ 

	$n = 10$	$n = 20$
$\mu_{(n)}$	1.5399	1.8675
$T_{(n)}(90)$	2.309	2.559
$T_{(n)}(95)$	2.568	2.799
$P_{(n)}(2.0)$	79.4%	63.1%

values are different. The ranking order is then always clear. The fact that identical values may occur in practice is a secondary phenomenon in the formulation of the present problem). We are therefore interested primarily in the behaviour of:

(2a) The extreme values  $X_{(1)}$  and  $X_{(n)}$ ,

and the range

(2b)  $R = X_{(n)} - X_{(1)}$ .

According to the new draft of the guidelines of the German Medical Council for guaranteeing quality in medical laboratories (1), values  $n = 10$  and  $n = 20$  are particularly important for determining the random sample range (length of series). Each of the quantities (1), (2a) and (2b) has its own probability distribution, with a corresponding expected value and a corresponding variance, and each pair of quantities (1) has a two-dimensional distribution with a corresponding correlation coefficient. The formula for the probability distribution of the extreme values is very simple, whereas that for the range is very complex. Conversely, the determination of the expected value and variance of extreme values generally presents numerical problems, whereas those of the range can be derived from the extreme values by simple calculation. For normal distributions, all these characteristic values are largely available (see e. g. l. c. (3–5)). For standard normal distribution and for the above-mentioned random sample range, the expected value  $\mu_{(n)}$  for extreme values, and the quantiles  $T_{(n)}(90)$  and  $T_{(n)}(95)$  for 90% and 95%, respectively, can be taken from table 1, which also gives the probability  $P_{(n)}(2.0)$  that a minimal deviation of two standard deviations will occur above (for  $X_{(n)}$ ) and below (for  $X_{(1)}$ ) the central value. It must be noted that the named quantities are all markedly larger than two (even for  $n = 5$ , the quantile for 90% is still somewhat larger than 2). For any normal distribution the quoted scale value must be multiplied by  $\sigma$  and referred to an expected value of  $\mu$ . For example:

A measurement system measures the serum glucose concentration without systematic errors, and with an analytical standard deviation of 5%. A control sample contains a glucose concentration of 5.56 mmol/l (100 mg/dl). In 10 control measurements in a stable measurement system the largest result exceeds the value

$$5.56 + 2.309 \times 0.28 = 6.21 \text{ (mmol/l)}$$

$$100 + 2.309 \times 5 = 111.545 \text{ (mg/dl)}$$

with a probability of 10%. The smallest result lies below the value

$$5.56 - 2.568 \times 0.28 = 4.84 \text{ (mmol/l)}$$

$$100 - 2.568 \times 5 = 87.16 \text{ (mg/dl)}$$

with a probability of 5%.

The probability that the value

$$5.56 + 2.0 \times 0.28 = 6.12 \text{ (mmol/l)}$$

$$100 + 2.0 \times 5 = 110 \text{ (mg/dl)}$$

will not be exceeded in 10 measurements is 79.4%; for 20 measurements the probability is 63.1%, assuming each time that the measurement system is stable.

From the above, the following statements can be made for the range: The expected value is

$$\begin{aligned} \mu(R) &= \mu(X_{(n)}) - \mu(X_{(1)}) = 2\mu_{(n)}\sigma \\ &= 3.078\sigma \text{ for } n = 10, \\ &\text{and } 3.735\sigma \text{ for } n = 20 \end{aligned}$$

For such small values of  $n$ ,  $X_{(n)}$  and  $X_{(1)}$  display a strong interdependence, which is why the probability distribution of  $R$  must be calculated according to a complex formula. The relevant values can be taken from the literature, in particular from l. c. (5). For a normal distribution the quantile  $T_R(90)$  and  $T_R(95)$  for 90% and 95% respectively, i. e. the values that are exceeded with a probability of 10% and 5% respectively, are:

$$\begin{aligned} n = 10: T_R(90) &= 4.13\sigma; T_R(95) = 4.47\sigma \\ n = 20: T_R(90) &= 4.69\sigma; T_R(95) = 5.01\sigma \end{aligned}$$

On the other hand, for  $n = 10$ ,  $R$  exceeds the value of  $4\sigma$  with a probability of about 13% (c. f. for  $n = 9$  the probability is about 10%, for  $n = 8$  about 9%).

Before reaching the obvious conclusion, these values should be compared with the corresponding values of suitably chosen  $Q$ -distributions (see Appendix). Table 2 shows the quantiles of table 1 (a) together with the corresponding values for the  $Q_3$ -distribution (b) and the  $Q_4$ -distribution (c). For the  $Q_3$ -distribution the interval of possible values is three times the standard deviation on

**Tab. 2** The quantiles from table 1, with those for the  $Q_3$  and the  $Q_4$ -distribution

	$T_{(n)}(90)$	$T_{(n)}(95)$
$n = 10$		
a	2.309	2.568
b	2.140	2.289
c	2.170	2.343
$n = 20$		
a	2.559	2.799
b	2.284	2.410
c	2.337	2.482

either side of the expected value; this must be taken as approximating to the smallest significant value of  $k$  for practical purposes. For the  $Q_4$ -distribution this value is  $\sqrt{11}$ -fold (see Appendix (9)). It is to be expected that these quantiles will be smaller than for a normal distribution, but it can be seen that even for  $k = 3$  the quantile for 90% is still larger than 2. The expected value for  $X_{(10)}$  was calculated for the case of  $Q_3$ :  $\mu_{(10)} = 1.51$ , i. e.  $\mu_{(R)} = 3.02$  (in non-standardized  $Q$ -distributions, just as in normal distributions, these numbers must still be multiplied by  $\sigma$ ).

Furthermore, the probability with which  $R$  exceeds the value 4 (or  $4\sigma$ ) was calculated for  $k$ -values between 2 and 5. These probabilities range from 5% to 9%.

### Control Limits and Stability of a Measurement System

Assuming a normal distribution or a similar distribution of the measured values, and an expected value  $\mu$  and standard deviation  $\sigma$  from the measurement system, let the set of the individual measurement values (which can be considered as compatible with the precondition of stability) be expressed in the form:

$$\mu - c_n\sigma \leq x \leq \mu + c_n\sigma$$

Then from the foregoing considerations and from table 1, it is clear that this factor  $c$  for the sizes of the random samples in question ( $n$  = number of control measurements that must be considered) must be at least 2. Thus the permissible interval for the results of control sample measurements is closely connected with the permissible interval for the systematic and for the random deviations of measurements. If, for the given component and target value,

$\mu_0$  is the smallest permissible value for the expected value from the measurement system,

$\mu_1$  is the largest permissible value for the expected value from the measurement system,

$\sigma_m$  is the maximal permissible standard deviation,

then, according to the above statement, the permissible interval for individual values must at least include the interval

$$(4) \quad \mu_0 - 2\sigma_m \leq x \leq \mu_1 + 2\sigma_m.$$

Therefore, if the distance of  $\mu_1$  or  $\mu_0$  from the target value has the same order of magnitude as  $\sigma_m$ , then the general span of the permissible set interval for individual measurements must be in the order of at least  $6\sigma_m$ . The consequences of this are shown by the following examples. To start with, it should be noted that *long-term* trends in the operational condition of the measurement system can only be detected by comparison of sequential control cycles. Such a comparison is, however,

not stipulated in the guidelines. It follows that *short-term* changes occurring *during* a control cycle can only be detected, and assessed for their compatibility with the requirements of a stable measurement system, by the investigation of extreme values.

### Example 1

Let the expected value from the measurement system under consideration lie somewhat above the smallest value of  $\mu_0$  that is still permissible; the standard deviation can be assigned any permissible value. A measurement value is obtained that is far greater than  $\mu_1$ , but is still within the control limits; this is acceptable according to the above rules. Such an occurrence is, however, very improbable in the type of measurement system under consideration, and can therefore be taken as an indication of a sudden change in the measurement system, or an error in the performance of the measurement. In either case, the measurement of patient samples will be affected in the same period. If this change or error appears only on some days during the control cycle, then the mean value and standard deviation will still meet the specified requirements.

### Example 2

Let the expected value from the measurement system be approximately equal to the target value, i. e. approximately half way between  $\mu_0$  and  $\mu_1$ ; let  $\sigma$  be much smaller than  $\sigma_m$ , e. g. about  $\sigma_m/2$ . For the reasons stated above, a control sample result that lies far below  $\mu_0$  or far above  $\mu_1$  must already be regarded as suspect, even if it still lies within the limits (4). Neither the mean value nor the standard deviation for the control cycle in question give any cause for alarm when such extreme values occur.

In my opinion, such examples clearly show that if the regulations (1) are to fulfil their purpose completely, an additional evaluation procedure is imperative for the examination of the values of individual control samples during each control cycle. Since opposition to the general use of quality control cards apparently precludes the introduction of the necessary prescriptions (crucial factors in this opposition may be avoidance of cost and waste, or the practicability of the procedure), at least a simple, easily performed and as far as possible non-expensive additional procedure should be introduced. It would seem that this requirement could be met by an evaluation of consecutive ranges.

As a rule, a control cycle is performed over a period of 20 working days, i. e. it consists of a series of 20 control sample measurements. For determination of the range, a random sample size of about  $n = 8$  appears to be optimal for many purposes (see l. c. (2)). The obvious procedure is therefore to divide the series of 20 results into

two series of 10 each (this applies equally well for other cycle lengths). According to the above specifications, the range then has a maximal permissible value in the order of 4 times the standard deviation; this value therefore forms the basis of the new formulation (1). As already explained (see end of section "Extreme Values and Range"), this specification is adequate for other, as well as normal distributions.

### Description of the control procedure

At the beginning of each control cycle the absolute difference between the two first control measurements is evaluated for the case of  $n = 10$ , assuming that this is the chosen size of random samples. It is then established whether the third control measurement lies within the range determined so far, i.e. it is determined whether the range is unchanged or not. If not, then the new range is determined and evaluated. This procedure is repeated for the next control measurement, and so on. If all the first 10 control measurements fulfil the criterion for  $n = 10$ , the procedure is started again with control measurements number 11 and 12. If the criterion is not fulfilled for any of the sequential control results, then appropriate measures must be taken (see l.c. (1)).

Let it be emphasised once again that such a procedure can only be expected to reveal short-term, relatively large changes in the operational condition of the measurement system or the performance of the measurements, but for the time being these are the changes of greatest consequence. More complex procedures are necessary for the detection of more complex or longer-lasting phenomena.

## Discussion

### Characterisation of the suggested procedure

#### *Transfer to more than two reported values*

The initial idea was to transfer the concept of "critical difference" between two identically distributed random variables  $X_1$  and  $X_2$ , namely

$$(5a) \quad c_{1-\alpha}\sigma = [|X_1 - X_2|]_{1-\alpha},$$

to more than two reported values. Formulation (5a) expresses the absolute difference, e.g. between an initial measurement and a subsequent measurement on a patient, which is not exceeded with a probability of  $1-\alpha$ . If it is exceeded in the measurement of a control sample, this is an indication that in the meantime an "alteration" may have occurred in the measurement system. For any normal distribution and  $\alpha = 0.10$ , the factor  $c_{1-\alpha}$  has the value  $1.645\sqrt{2} = 2.326$ . Obviously, the absolute difference is identical with the range of the two  $X$ -values. For the general case, the "critical maximal absolute dif-

ference" of all the pairs of measurement values (i.e. the range  $R$ ) should therefore be taken as the basis:

$$(5b) \quad C_{1-\alpha}\sigma = [R]_{1-\alpha}.$$

This is the value of the range that is not exceeded with a probability of  $1-\alpha$ . In addition, the critical value  $C$  still depends on the number of measurement values, and for a normal distribution with  $n = 10$  and  $\alpha = 0.10$ , it has the value of 4.13, which was given earlier in section 3 for  $T_R(90)$ .

### *Relationship to quality control cards*

If the actual imprecision of the measurement procedure, as realised in the measurement system, is taken as a basis, and the final range is recorded at the end of each series of control measurements, then this is identical with the plan of a "range card" (see e.g. l.c. (6)). In the suggested procedure there are two departures. First, the critical event must be taken into account as soon as it occurs, and its consequences must not be postponed until after the conclusion of the measurement series. Second, an adequate estimate of the real standard deviation does not need to be made and applied beforehand (this would confer relative maximal effectiveness on the procedure), but rather, under certain conditions, even the maximal permissible imprecision may be used (of course, this can considerably reduce the effectiveness).

### *Relationship to outlier procedures*

The suggested procedure can, at least when the actual imprecision is used, be considered as an outlier test, and it is presented as such in l.c. (7). In contrast to all other-wise normal procedures, criterion (5b) is applied sequentially and its introduction is not delayed until the end of the sampling procedure. However, it is not attempted to set up the sequential procedure as a multiple test, in that a critical value for the case of  $n = 2$  is inserted for the first two control results, followed by a new critical value for  $n = 3$ , and so on. This would make the procedure too complicated for practical purposes, and furthermore the problem of the choice of sequential  $\alpha$ -values would first have to be solved.

### Choice of confidence probability

A value of 90% is suggested here for the confidence probability. A probability of 95% (or 5% for the test level) would otherwise normally be used for statistical evaluation. Apart from the fact that, almost always, the use of this value is only a convention, it is also relevant to the testing of null hypothesis. In the present case, it is a matter of striking a reasonable balance between the early detection of serious cases, and a minimal number of random alarms. For this reason a confidence probability of 90% (i.e. a test level or error probability of the first kind of 10%) would appear to represent an ac-

ceptable risk. Incidentally, before the suggested procedure is introduced, it should undergo suitable testing.

### Use of estimated values

The present work is based on the theoretical model of probability. In practice, this is satisfactory for stability control, since the individual measurement results correspond exactly to the random values under consideration. The mean value does not feature in the suggested procedure, but the theoretical standard deviation  $\sigma$  must be replaced with an estimated value or at least a permissible value. If the inserted value accurately represents the theoretical value, then the results will also correspond to theory. Otherwise, there will be a decrease of sensitivity or an increase in the number of random warnings.

### Monitoring of the mean value and standard deviations

A completely different problem is encountered in the establishment of the control limits for the systematic and random deviations in measurement results. This necessitates the inclusion of quite different criteria, which are related e.g. to the number of correct/false decisions determined by clinically important decision limits (see l.c. (8)).

### The numerical establishment of the criterion

For two practical reasons, the maximal permissible value for the range was established as exactly 4 times the currently used standard deviation. First, experience shows that such smooth values are much easier to handle without error than fractal broken values. Second, as in other comparable situations, such a smooth value makes allowance for the absence of a figure for the actual distribution. There may also be differences between the standard deviation used and that which ought to be used. It must therefore be accepted that the confidence level is only a "fuzzy" value, i.e. it can only be given as "approximately" 90%. Consideration of  $Q_k$ -distributions makes it clear that at least ignorance of the actual type of distribution of the measured values does not have very serious consequences. In other words, the factor of 4 can be considered as a kind of "balanced value" for numerous possible situations.

### Appendix

It is not intended and neither is it possible to deal with all the important statistical aspects of the  $Q$ -distributions. In introducing this family of distributions, my only intention is to show that the suggested numerical treatments of extreme values and ranges are not necessarily restricted to a normal distribution, and that these quantities do not acquire fundamentally different values when certain deviations from the normal distribution occur. The most important type of deviation considered was that in which the scale of values on either side of the mean is not theoretically infinite, but has upper and lower limits. This would appear to correspond to

reality. On the other hand, the property of a symmetrical distribution around the expected value (mean) was retained. This may sometimes be unrealistic, in particular when dealing with extreme scale values, e.g. near to the detection limit. But such situations do not normally arise with control samples.

### Establishment of the type of distribution

1) In a real measurement system performing multiple measurements on identical samples, one does not expect arbitrarily large deviations from the mean value, but rather that the measurement values will fluctuate within a lower and an upper limit. These limits depend on the component being measured and the properties of the measurement procedure, and they should not be considered as rigidly fixed. The situation is best illustrated by comparing it to the problem of how many handkerchiefs will fit into a suitcase (see account in l.c. (9)). It is clearly not possible to fit an infinite number of handkerchiefs into a given suitcase. However, suppose the suitcase contains 1000 handkerchiefs and appears full, it will always be possible to cram in a few more, but certainly not another 1000. Similarly, in a measurement system, in the measurement of any given quantity, values will occur with a deviation from the mean up to a certain multiple of the standard deviation. However, the measurement system can basically "generate" a somewhat larger deviation, but this can never be twice the latter multiple.

Such a situation is adequately represented mathematically by probability density functions, which, below the value  $a$  and above the value  $b > a$ , are equal to 0. This does not mean that the occurrence of measurement values somewhat outside these limits is absolutely impossible!

The simplest probability density functions of this kind are those that are proportional to an expression of the form

$$(x-a)^k (b-x)^l$$

with an appropriately chosen proportionality factor for the interval

$$a \leq x \leq b$$

Outside the limits of this interval, the probability density function should be set at 0.

2) For these probability density functions to rapidly approach 0 at the end points  $a$  and  $b$ , the first derivation of these points must also approximate to 0, i.e.

$$k > 1, l > 1$$

3) All these probability density functions then increase continually from  $a$  (where they are equal to 0) to a maximum. They then decrease again until they attain a value of 0 at  $b$ . If the maximal value lies exactly in the middle of the interval  $(a, b)$ , then the following applies:

$$k = 1$$

In this case the probability density function is symmetrical to the mid-point  $(a + b)/2$  and can be written in the form:

$$(6) \quad \begin{aligned} f_k(x|\mu, A) &= C \cdot (A^2 - (x - \mu)^2)^k \\ &\text{for } \mu - A \leq x \leq \mu + A \\ f_k(x|\mu, A) &= 0 \\ &\text{for all other cases,} \\ \mu &= \frac{1}{2}(a + b), A = \frac{1}{2}(b - a) \end{aligned}$$

4) By the transformation

$$y = \frac{1}{2}(1 + (x - \mu) \cdot A^{-1}), x = \mu + A(2y - 1)$$

(6) can be converted to the form

$$(7) \quad b_{k+1, k+1}(x) = C' \cdot y^k (1 - y)^k, \\ \text{where } 0 \leq y \leq 1.$$

This is the probability density function of a "beta distribution of the first kind", with the parameters  $k + 1, k + 1$ . This immediately yields a value for  $C$ , which must be chosen in such a way that the integral of (6) over the interval of  $\mu - A$  to  $\mu + A$  is exactly equal to 1. First,  $C'$  can be expressed as

$$C' = \Gamma(2k + 2) \Gamma^{-2}(k + 1),$$

from which it follows that:

$$C = \Gamma(2k + 2) \Gamma^{-2}(k + 1) \cdot (2A)^{-2k-1}.$$

( $\Gamma(u)$  is the gamma function, which for integers takes on the value of  $\Gamma(u) = (u - 1)!$ )

We have named this distribution a "Q-distribution", because its definition is based a quadratic expression.

#### Remarks

The type of distribution described here is known in the literature as type I, or (in the form of (6)) as type II in the classification of *K. Pearson* (see e. g. l. c. (10)). In its transformed version as a beta-type distribution, it plays a role in various theoretical problems, but it does not appear to have been investigated for its applicability to actual experimental results.

#### Definition of Q-distributions

The class of Q-distributions depends on three parameters:

- Central value  $\mu$
- Scale width  $A > 0$
- Exponent  $k > -1$ , or  $k > 1$ .

The complete probability density function is represented by:

$$f_k(x|\mu, A) = \Gamma(2k + 2) \Gamma^{-2}(k + 1) (2A)^{-2k-1} [A^2 - (x - \mu)^2]^k$$

$$\text{for } \mu - A \leq x \leq \mu + A$$

$$f_k(x|\mu, A) = 0$$

for all other cases;

$\mu$  is the expected value.

The relationship between the width of the scale and the standard deviation is expressed as:

$$(8) \quad A = \sigma(2k + 3)^{1/2}$$

#### Standardized Q-distributions ( $Q_k$ -distributions)

Characterisation of a standardized Q-distribution requires only one parameter, namely an exponent  $k$ . The expected value is set at 0, and the standard deviation at 1. According to l. c. (8), this then gives the relationship between scale width and  $k$ :

$$(9) \quad A = (2k + 3)^{1/2}$$

The probability density function then appears as follows:

$$g_k(x) = \Gamma(2k + 2) \Gamma^{-2}(k + 1) 2^{-2k-1} (2k + 3)^{-k-1/2} (2k + 3 - x^2)^k$$

$$\text{for } |x| \leq (2k + 3)^{1/2},$$

$$g_k(x) = 0$$

$$\text{for all other cases.}$$

Example:  $k = 3$

$$(10) \quad g_3(x) = 7! (3!)^{-2} 2^{-7} \cdot 9^{-3-1/2} (9 - x^2)^3$$

$$= 35/96 \cdot 9^{-3} (9 - x^2)^3 = 35/96 (1 - x^2/9)^3$$

(see fig. 1)

Since the definition of the standardized Q-distribution depends on one parameter, namely the value of exponent  $k$ , it is referred to as a  $Q_k$ -distribution.

Like a normal distribution, any Q-distribution can also be standardised by the transformation

$$Y = (X - \mu)/\sigma.$$

This transforms the three-parametric family of Q-distributions into a one-parametric family, namely that of the  $Q_k$ -distributions. In contrast, there is only one standardized normal distribution.

The relationship of the  $Q_k$ -distributions to the standardized normal distribution is described by the following property:

If the value of the parameter  $k$  increases against  $\infty$ , then  $g_k(x)$  converges on the probability density function of the standardized normal distribution.

For  $k = 3$  the maximal difference between the density function of the standardized Q-distribution and that of the standardized normal distribution is at the most equal to

$$(2\pi)^{-1/2} - 35/96 = 0.03436,$$

and this maximal difference occurs at the point  $x = 0$ . In fig. 1 the scale of the ordinate has been chosen so that the difference at this point ( $x = 0$ ) can be clearly seen; for larger  $x$ -values, however, this difference is hardly recognisable. Nevertheless the two density functions are mathematically fundamentally different.

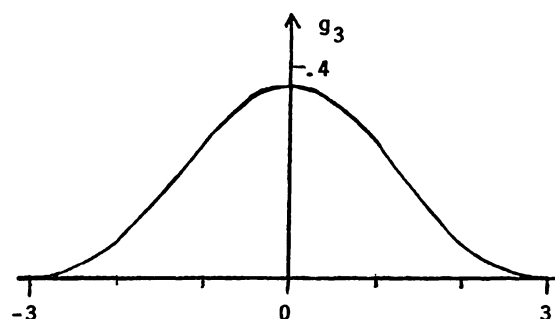


Fig. 1 Density function  $g_3(x)$  of the standardized Q-distribution for  $k = 3$  (for  $Q_3$ -distribution, see Appendix, equation (10)). For  $|x| > 3$  the density function is set at 0. In order to illustrate the difference between the normal and the Q-distribution more clearly, the scale of the ordinate is 6.25 times that of the abscissa.

#### References

1. Entwurf Richtlinien der Bundesärztekammer zur Qualitätssicherung in medizinischen Laboratorien (Stand: 27. 6. 95/13. 7. 95).
2. David HA. Order statistics. New York: J. Wiley & Sons, 1970.
3. Pearson ES, Hartley HO. Biometrika tables for statisticians Vol. I. 3rd ed. Cambridge: Cambridge UP, 1970.
4. Pearson ES, Hartley HO. Biometrika tables for statisticians Vol. II. 3rd ed. Cambridge: Cambridge UP, 1972.
5. Harter HL. Tables of range and studentized range. Ann Math Stat 1960; 31:1122-47.
6. Uhlmann W. Statistische Qualitätskontrolle. Stuttgart: Teubner Verlagsgesellschaft, 1966.
7. Barnett V, Lewis T. Outliers in statistical data. New York: J. Wiley & Sons, 1978.
8. Stamm D. The influence of the reliability of clinical chemical analysis on medical decisions. In: 31st International Congress of Pure and Applied Chemistry: Invited Lectures, Section 9: Clinical Chemistry; 1987; Sofia. Bulgarian Academy of Sciences, 148-170.
9. v. Weizsäcker CF. Zum Weltbild der Physik. 10. Auflage Stuttgart: Hirzel Verlag, 1963.
10. Ord JK. Pearson system of distributions. In: Encyclopedia of statistical sciences. New York: J. Wiley & Sons, Vol. 6, 1985.

Received October 31, 1995/May 9, 1996

Corresponding author: Erwin Hansert, Kraepelinstraße 4, D-80804 München, Germany

